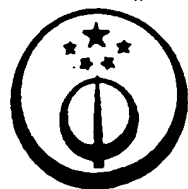


Pending Application

[19]中华人民共和国专利局

[11] 公开号 CN 1133450A



[12] 发明专利申请公开说明书

[21] 申请号 95104620.9

[51] Int. Cl⁶

G06F 3/023

[43] 公开日 1996 年 10 月 16 日

[22] 申请日 95.4.19

[30] 优先权

[32] 94.4.28 [33] CN[31] 94104738.5

[71] 申请人 黄宝珩

地址 310013 浙江省杭州市曙光路求是村
71-204

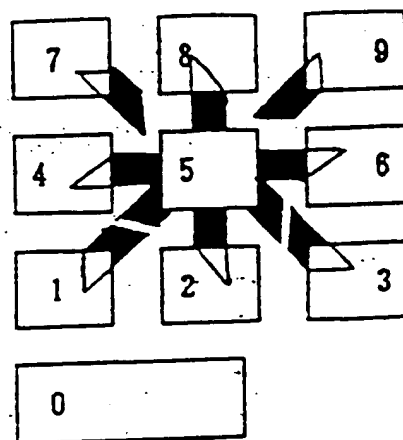
[72] 发明人 黄宝珩

权利要求书 1 页 说明书 10 页 附图页数 1 页

[54] 发明名称 汉字“笔画分子”型编码输入方法——简称“分子码”又称“米字码”

[57] 摘要

“米字码”属计算机全形码汉字输入法，采用了小键盘“米字笔画象形排序”法，四位数字码，半自动选字法，单手输入，特点（1）每键记符 < 2.5 个，难度系数 Dh 为常用码的 1/10。（2）米字键盘表，“米、口”字引导图，金木水火土等 10 部首，易记好用停用无妨，适合于老中青少，方言和 PC 进入家庭，重码比拼音少 70%，最多重字少 90%，且 90—95% 以上只重 2 字或 3 字，对国标一，二级字库的静态重码率各 < 25%，半自动选字法，可用软件等效再辅降重码 50—70%。



(BJ)第 1456 号

权 利 要 求 书

1. 一种汉字“笔画分子”型编码方法——简称“分子码”，又称“米字码”属计算机汉字输入技术领域的一种全形码，它包括：
A. 笔画分子概念， B. 汉字的米字定律， C. 米字码的三组份，
D. 十个数字，四位码，22含义符设定， E. 小键盘米字象形排序法组成代码，
F. 中心线取码法， G. 千位，百位数部首降重码调控，
H. “数”参与汉字编码， I. 半自动选字法， J. 米字码引导图，
本发明的特征是所述的笔画分子概念是指：任何汉字的各种笔画均可以看成是由一个以上的具有矢量特征的，相同的等腰梯形的笔画分子所组成。
2. 根据权利要求1所述的汉字的米字定律，其特征是指：全部汉字的笔画分子矢量可以概括为在平面上互差45度的七个不同矢量，可以用米字的七个端点的笔画分子矢量表示：[A]，[B]，[C]，[D]，[E]，[F]，[G]，
3. 根据权利要求1所述的三组份22个含义符，其特征是笔画分子码七个即：[A]横右，[B]竖上，[C]横左（含独横），[D]撇，[E]竖下及单竖，[F]捺，[G]点，部件码三个即：[H]口，[I]勾，[J]十，X，<，此三含义均称为“交”码，及特定部首码十个即：车，木，艹，彳，火，土，亻，亻，月，扌，
4. 根据权利要求1所述的十个数字码符，其特征是0-9十个数字可分别用作笔画分子码，部件码，及特定部首码共20个码的代码，代码的配置可以是按米字端点矢量旋转顺序（简称矢量序）即：
[A]为2，[B]3，[C]4，[D]5，[E]6，[F]7，[G]8，[H]0，[I]1，[J]9。
5. 根据权利要求1所述的十个数字码符，其特征是0-9十个数字可分别用作笔画分子码，部件码，及部首码共20个码的代码，代码的配置可以是按计算机小键盘的米字象形排列顺序配置代码（简称象形序）即：
[A]为6，[B]5，[C]4，[D]1，[E]2，[F]3，[G]7，[H]0，[I]8，[J]9。
6. 根据权利要求1所述的中心线取码法，其特征是基本规则按“中心线，首，尾，转”取码，即对任一汉字均可设定画一道以上的垂直中心线，与线相交的分体部件的首尾笔画即为取码对象，不足四码时再顺时针取码（遇右侧向勾则逆时针取码），取码中心线可有1-4条，以适应各种字型。
7. 根据权利要求1所述的降重码千，百位数部首调控，其特征是令十个部首赋予特定的代码，可以是米字矢量序即：
车 61，木 76，艹 27，彳 37，火 48，土 47，亻 67，亻 77，月 07，扌 17，
8. 根据权利要求1所述的降重码千，百位数部首调控，其特征是令十个部首赋予特定的代码，也可以是小键盘米字象形序：
车 23，木 35，艹 63，彳 53，火 45，土 43，亻 23，亻 33，月 03，扌 83，
9. 根据权利要求1所述的“数”参与汉字编码，其特征是对任一汉字结构中相同的“口”，“点”，“交”笔画部件的代码可用数字跟后表示其数目，
10. 根据权利要求1所述的半自动选字，其特征是在屏幕提示重字栏内“设置自动步进循环的”方框选择背景”，再按任意键选字。
11. 根据权利要求1所述的米字码引导图，即指“米”，“口”，“克”三个字图及其编码，和口字的少一笔，及少两笔的8个单元码，可有效快速学用掌握米字码。

汉字笔画分子"型编码输入方法——简称"分子码"又称"米字码"

汉字"笔画分子"型编码输入方法——简称"分子码",因与"汉字的米字定律"和米字引导图有关,又称"米字码",属计算机汉字输入技术领域的一种全形码输入法。有关汉字输入技术的背景文件,著作:

- | | | |
|-------------------|-------|-----------------|
| [1] 汉字终端技术入门 | (周关兴) | 人民邮电出版社 1992.2 |
| [2] 计算机常用汉字输入方法 | (王盼卿) | 电子工业出版社 1992.10 |
| [3] 启宏全息码汉字编码输入系统 | (萧启宏) | 电子工业出版社 1993.3 |
| [4] 计算机文字处理与信息管理 | (吴良占) | 杭州大学出版社 1990.0 |
| [5] 通用汉字输入方法及排版系统 | (邵平凡) | 浙大科海公司 1994.2 |
| [6] 汉字信息处理 | (陈增武) | 贵州人民出版社 1988.5 |
| [7] 计算机应用基础教程 | (周 苏) | 天津科技出版社 1993.4 |
| [8] 自然码汉字输入法 | (周志农) | 浙江科技出版社 1993.1 |
| [9] 表音码汉字输入法 | (沈克成) | 浙江科技出版社 1994.7 |

发明一种易学易用,好记难忘,中外老少,普及适用的"小键盘米字象形键盘编码"的汉字输入方法,以适应面临21世纪前后计算机将普遍进入家庭的需要。计算机这一人类文明的生产工具和生活高质量的标志,现已获全球共识。在进入2000年的十年期间,在高效率办公,高信息传输,存储,高速运算,大范围联网,录取,显示,等方面功效卓著,加上微电子,光纤通信,多媒体的进展,推动了PC486,奔腾PENTIUM PC586,强力POWER PC604,以及95年刚推出的INTEL P6芯片的PC686等多种新的PC机型的大发展,大竞争,大跌价和随之而来的必然的计算机大普及浪潮,计算机将像电话机一样多地进入到家庭。若使用汉字信息处理者的年龄范围扩展至[10--70-80岁],则市场前景,社会需求量总在数亿人以上。

"计算机汉字输入方法"一直是汉字文化圈内计算机普及的一个瓶颈。汉字输入编码方法有:音码,音形码,和形码这三种类型。国内外历时几十年,方案数百种,迄今仍未解决好,仍处普及的瓶颈态势。前5-10年发明的各种汉字输入方法在解决计算机排版,机关文秘等方面虽功不可灭,但一般中老年人仍望而生畏,年青人除专业者外,多数宁慢试打拼音,也不想强背,难记易忘的各种键盘表。试想人们为解决拼音输入重音字多,方言多,难普及的缺点,已奋斗了三十余年,提出了三百多种详硕方案迄今仍是多数宁慢试打拼音,则足以说明现行汉字输入方法弊端之严重不理想。关键是要解决输入编码的易学,好记,和难忘的这个课题。本发明的目即为克服现有技术之不足,解决计算机普及的这个汉字输入瓶颈,以迎接21世纪信息时代的到来。

对现有技术的评估,作者认为对"汉字输入法"有三个问题亟待取得共识:

1. 汉字编码理论未获重视,多数人埋头发明实践精神感人,然而理论升华,理论定位功亏一篑。现代复杂高科技之理想突破进展,无不是在理论分析指导下取得的。
2. 评估鉴定时看重用户操作比赛结果,印象评语,而量化指标项目不足,无先进性定位,用户难辨东西南北。
3. 高速盲打成为重要的,甚至成为唯一受重视的指标,形成误导。许多编码方案特设一级,二级,三级"简码",要求操作者"熟练掌握"(背车)500,1000,2000个"简码"。当年机械打字亦不

过背记三千。汉字输入方案须针对新的应用对象需求,其评估方法亦应有新的侧重,这就是要低难度系数(易学好记),宽的人口,年龄,民族,地域范围,输入速度可降为第二,三位,甚至以不影响拟稿即可。作者着重使应用与理论分析挂钩,明确先进性定位的量化指标,并对米字码与已有技术进行评估对比。

现行汉字输入法多侧重于解决“输入速度”和“重码率”两项指标,目标针对专业操作人员,而输入方法之难易,则不在严格定量评估之列,致使普及发展滞缓。而把计算机当作工具的非专业操作者(这类将占八九成之多),对“难度”最为重视,也是电脑进入家庭的障碍,故作者明确提出平均难度系数D这一概念:定义D为:直观(硬)难度系数Dh及与操作细则,图表记忆方法,取码方法等有关的(软)难度系数Ds的乘积,即: $D = D_h \times D_s$ (4)

硬难度系数Dh可定义为平均每打一个英文字(Word)或汉字的操作复杂程度(操作难度)和所需的记忆量(记忆难度)。键盘大,击键次数多,单键(码元)含义数多,需背需记的编码数多所达几百,上千,几千,其难度系数则高,反之则低。难度系数是没有单位的量化指标,但可相对达几百,上千,几千,其难度系数则高,反之则低。难度系数是没有单位的量化指标,但可相对比较,最好的比较对象是英文,它不学自会,以英文为基准的硬难度系数Dh值的大小,一般可以反映某种输入方法的主要难度,尤其是键盘的操作,记忆难度。例如:平均每打一个英文字的难度系数(Dh)为: $D_h = 26 \times (5.2-5.8) = 143$ (每英文字平均为5.2-5.8个字母) (5)

公式: $D_h = A \times B \times K \times N \times L$

K = 码符总数(即输入码需动用的总键数,称码元数)

N = 平均每键的含义数 L = 平均每字击键数 (暂定非等长码为L+1)

A = 编码体制加权系数 (暂定音码,形码均为1,音形码复合体制为1.4)

B = 码符加权系数 (暂定:码符可读为1,有序为1,反之为1.2-1.4)

直观难度系数 Dh 能反映键盘记忆难度,而 Ds 反映取码难度,操作规则记忆难度等。Dh虽不全面反映难度,却能以数字相互比较主要难度部分,有比较鉴别才能促使需要和改进的方向发展。因而颇有参考价值。不同的汉字结构分析所得出的编码方法不同,它的A,B,K,N,L等参数各异,因此Dh值也可作为改进汉字输入方法的计算工具之一。此外汉字不同于英文,除硬难度Dh以外,例如取码规则的多少,即便条文数目相等,其难度差别也大,上手感觉如何,停用是否易忘,这些不好量化的难度,称为软难度系数Ds, Ds值可大面积统计评估得出。硬难度系的实例计算及列表对照。

表1. 为国内20余种著名汉字输入法,按Dh公式初步计算的结果。(附注重码重字等性能)

| 音 码 | Dh 值 | 音形码 | Dh值 | 形 码 | Dh值 |
|-----------|------------|--------|------|-------------|------|
| 英文 | 143 | J 码 | 344 | ** 米 字 码 | 92 |
| *****全 拼 | 300 | S 码 | 735 | ***** W 笔 | 96 |
| *** 人 大 码 | 342(3700字) | HY 码 | 1062 | ^### B 笔 | >151 |
| 通信拼音 | 360 | # Z 码 | 1656 | ^***** 四角号码 | 205 |
| *****双拼双音 | 380 | && C 码 | 4650 | 澳洲## 基本笔画 | 235 |
| | | | | HP 码 | 806 |
| | | | | 美~\$ 王安三角 | 918 |
| | | | | D 码 | 991 |
| | | | | L 码 | 1248 |
| | | | | ^^ W 码 | 1423 |
| | | | | 台^& 仓吉字母 | 1567 |

注1: *静态重码率P<22-25%(3755字库) *** >80-96%

+ 最多重字<4-6字 ++<15字 +++<50字 ++++<110字

非等长码最多 5位 ## 7位 ### 7-9位

^ 进入部分系统汉卡 ^^ 几乎全进 \$ 6位等长

& 可打1.6万字 && 5. 万字 其它6763字

九是要
三至以
三字母

三人员
三的非
三者明
三亿方

变)和
数多
相对
般可
文字

1)

Dh虽
进的
(N,
文,
四何
硬

以下对七种实例计算Dh值:

五笔字型:王(永明教授)码有:

成字根103个,键名根25个,特殊根72个(不在国标字符集内),共200个,(王码教材只言基本根130个,引自资料[5]P.4.) 交叉识别码 $3 \times 5 = 15$ 个,降重码L,容错码U,F,D,共4个,高频字码25个,Z键三功能,共计247个码符合义量。

$$K = 26. \quad L = 4(\text{单字码长}) \quad N = (200 + 3 \times 5 + 7 + 25) / 26 = 9.5$$

$$\text{约40\%字根难读,加权}(1.2). \text{字根有序加权}(1.2) \quad B = 1.2 \times 1.2 = 1.44$$

$$\text{得: } Dh = 26 \times 9.5 \times 4 \times 1.44 = 1423$$

资料[5]提出:训练要求背熟王码一级简码(1键码)25个高频字之外,牢记二级简码(2键码)588个,和熟练掌握三级简码(3键码)2250个。一,二,三级简码的利用率要达到95%以上,通过训练使达到打速为800字/10分以上,然此处Dh值当中尚未计及或尚未反映其背记难度。

2. 见字识码法——上海支(秉彝博士)码有:

码元—26个,常用字元12个,字元标记50个,关系字X,Y,Z,R,H,Q,N,O,8个,高频字26个。

$$K = 26 \quad N = (12 + 50 + 8 + 26) / 26 = 3.7 \quad L = 4 \quad B = 1.2 \times 1.2 = 1.44$$

$$\text{得: } Dh = 26 \times 3.7 \times 4 \times 1.44 = 554$$

3. 通信拼音字——总参通信部杜(牧平)码有:

H,CH,SH双声代码V,W,Y三个,汉拼一通拼代码26个,一码双韵7个,声调代码24个。

高频字18个。零声母(OB,EL,OX,EX) 2x4个,即:

$$K = 26 \quad N = (3 + 26 + 7 + 2 \times 4 + 24 + 18) / 26 = 3.3 \quad L = 2.5 \quad B = 1.2$$

$$\text{得: } Dh = 26 \times 3.3 \times (2.5 + 1) \times 1.2 = 360.$$

4. 自然码——周志农

音码——码元26,代表全拼38音。形码码元30个,代表部首组件168个,字型码4个,笔画码8个(多数是用标点符号当作代码)。

$$K = 30 \quad N = (38 + 168 + 4 + 8) / 30 = 7.3 \quad L = 2-4 \quad A = 1.4 \quad B = 1.2$$

$$\text{得 } Dh = 30 \times 7.3 \times (3.5 + 1) \times 1.4 \times 1.2 = 1656$$

5. 全息码——蓄(启宏)码有:

码元——36个,形码(字元102,部首45,部件30,笔画8,规则键2,共187个,转注字元54个)

音码(一级字元329个,二级字元230个)共559个

$$K = 36 \quad N = (187 + 54 + 329 + 230) / 36 = 22 \quad L = 2.46 \quad A = 1.4 \quad B = 1.2$$

$$\text{得 } Dh = 36 \times 22 \times (2.46 + 1) \times 1.4 \times 1.2 = 4650$$

6. 王安三角——王安电脑三角编码(笑)

码元——10个, 306个基本笔形符号其代码共99个(双位代码01——99), 码长6。

$$k = 99 \quad 2N = 306 / 99 = 3.09 \quad L = 6 \text{ (3个双位码)} \quad \text{得 } Dh = 99 \times 3 \times 3.09 / 2 = 918$$

7. 米字码——黄(宝珩高工)码有:

笔画码7个[(左,右)横,(上,下)竖,撇,捺,], 部件码3个[口,勾,交(十X<)](含义共为5个), 千位数调控码9个:[(月03, 扌83, 扌23, 扌33, 扌43, 扌53, 扌63),

(木35, 火45, 扌65)] 其中百位调控码2个为[3],[5]

$$K = 10 \quad N = (7 + 5 + 9 + 2) / 10 = 2.3 \quad L = 4 \quad A = 1 \quad B = 1$$

$$\text{因笔画分子名称均可读,矢量等间隔有序,代码数字有序. 得: } Dh = 10 \times 2.3 \times 4 = 92.$$

汉字结构分析繁简差异会影响编码的复杂程度。定义S值为编码组成的成份：计有部首、笔画、部件、字根、字元、字型、字义等6-7种。每组分当中又包含若干个意义例：字型有上下型、左右型等，合计即称为总含义数H。H即为使用代码的总需方（方案不同H有5-800个之差异），若H值小，必促使加大码长L，才能描述并区别众多的汉字，否则重码太多。而码长L加大引起击键次数增多，输入难度增大，若既要降低码长L而又要重码少，则必须增多码元（码符），这就加大记忆量也使输入难度增大。根据编码基本原理可推出静态重码率Pc与编码参数的关系式：

$$Pc = H \times N / [Qc \times S \times K^L] \quad (2)$$

$$Qc = H \times N / [Pc \times S \times K^{L-1}] \quad (3)$$

式中：H 为汉字（库）总数

S 编码组成的成分数

N 平均每码元（码符或每键）所代表的含义数， $N = H/K$

K 编码所使用的总码元（码符）数

L 码长（平均每字击键次数）

Qc 即为各种汉字编码方法的优值系数，反映既定字库量，既定重码率情况下与编码复杂性成反比的参数，Qc越高越好。以相同的H=6763字代入Pc公式，算得下述三种不同的编码：

八笔码 K1=8码元 N=2 S1=2 L1=6 Pc1=0.026/Qc1（实际L=7-9的非等长码）

五笔码 K2=25码元 N=10 S2=4 L2=4 Pc2=0.043/Qc2（实际为2,3,4位有限元等长码）

米字码 K3=10码元 N=2.3 S3=3 L3=4 Pc3=0.54/Qc3

如代入其实际静态重码率：Pc1=0.04 Pc2=0.08 *Pc3=0.4(6763字) *Pc3'=0.25(3755字)

则得：Qc1=0.65 Qc2=0.54 Qc3=1.35 Qc3'=1.15

注：(1)Qc公式只对等长码准确，故Q1, Q2仅为近似值。(2)*Pc 为未计及软件降重数据。

以下按照中国中文信息研究 汉字编码专业委员会1983年“汉字编码评测试行规则”中所列重要指标进行的计算比较：

编码效率 $E = H(\text{汉字}) / [L \times \log_2 K]$ (4)

式中：H(汉字)为汉字平均每一码元的理论比特数（码元理论信息量）约9.6比特

L为平均每一汉字编码所用的码元数，即码长。

k为编码总码元（码符）数

由公式算出26英文字母的四码的编码效率 $E_e = 9.6 / 4 \times \log_2 26 = 0.51$

26英文字母的三码（例音形码）的编码效率 $E_e' = 9.6 / 3 \times \log_2 26 = 0.68$

10数字的四位码的编码效率 $E_d = 9.6 / 4 \times \log_2 10 = 0.72$

故 $E_d/E_e = 1.4$ 得出四位数字等长编码效率优于四键英文码1.4倍，亦优于三键英文码。

键入速度 $S = 60 / [(a+b \log_2 K) \times L]$ 字/分钟 (5)

式中：a, b系数由实测得出。若系数a, b相等则： $[S_d(\text{数字码})/S_e(\text{英文码})] > 1$ 数字键入快些。

静态重码率 $Pc = \text{重码数} C / W(\text{字库总数})$ (6)

式中：C = 重码字数 Z - 重码组数 n

(7)

综合Dh, Qc, Pc, Ed, Sd计算分析结果：

(1)Qc值越高确实反映了编码越趋简单。

(2)米字码以Dh, Qc值先进性定位均居优（相差>2-3倍）

(3)米字码Pc居中偏低（*数据未计及软件降重效果），软件降重后接近优级。

有部首,
有上下
90个之
5码长
多码元
2c与重

编码重
5编码:

一长码

7533)

5" 11

码

快息

对编码类型的综合评述:音码难度低,上手快是优点.缺点是:1.重码,重字过多,难于承
2.受地域方言障碍多,难于推广.尤其在中老年市场小.但多数青年(非专业人员),凭
拼音功底,宁慢试打拼音,也不学难记的现行汉字输入法.此点正说明汉字输入现状不
理想.音形码以复体制降低重码,改进音码.代价需增加记忆,然仍未跳出音码的地域,方
障碍.输入速率可高于形码,Dh值约为音码的5-10倍,因是两种体制的混合产物,故Dh.
值相形属最高.然而近几年所推出的音形码发明甚多,除表中所列五种之外还有:蓄码,
码,四角音形码一等.估计在汉字输入无“难度量化”指标的状况下,参照竞争对象又是
值本来就很高的五笔字型码(王码),则还将有更多的音形码面世.如果是以“通用,
及,高速”为目标,适应计算机进入2-3亿家庭的需要,则发明参照对象应是英文.即Dh
应低于200,使之与英文的Dh值 143接近,故就此而论音形码相去尚远.

形码体制花样多,难度差异大,然而创新改进潜力极大,又符合东亚汉字文化圈使用汉
国关于振兴汉字,统一常用汉字,的呼声与要求.形码较能符合计算机进入家庭,低难度
数的要求.形码最有可能追赶英文.表1所列十种形码方案,多数有十年以上历史.所知
年形码发明不多.形码的Dh值大小相差20倍.指标性能亦大有差别.由此反映出形码体
的灵活多样性,改进余地潜力很大.发掘中华文化遗产汉字宝库,试与英文比高低,形码
望极大.本发明“米字码”即是一个在Dh值较低(92)情况下获得中低重码率的成果的例
子.成果的取得是基于对现有技术,码特点的剖析认识和某些概念的更新,这些特点是:

- 1)推广最多的在大陆是五笔字型,台湾是仓颉字母,均为形码,决非偶然.说明推广音码
或音形码的困难大于形码.
- 2)过高的重码率(80-90%)不好应用.四角号码除重字多外(有多达40字),取码规则,附则
亦繁,属淘汰之列.
- 3)五键五笔画也是重字多,拆字繁.故每字击键不宜过四,否则Dh 值虽低亦难应用.
- 4)非等长码虽可降低平均击键次数,但多键字易出错,纠错亦难.操作不易规范化,不易
被接受.应慎用非等长码.最长为 9位的汉字笔形编码法(李码Dh=151)及基本笔画笔
顺法(芦码 Dh=235),属于这类.
- 5)现行形码降低重码,简化取码规则采用的办法明显趋向两个极端:多数过份依赖拘
泥于汉字的200个偏旁部首,致使Dh值起点很高(王码Dh=1423).少数则完全抛开部首
不用,备受重码困扰,Dh值虽低亦无用.
- 6)此外还有采用36键,45键方案,以扩增码元图降重的方案等,也因复杂不够理想.

一种汉字“笔画分子”型编码方法—简称“分子码”,又称“米字码”属计算机汉字输入技术
领域的一种全形码.米字码的定义为:一种运用笔画分子的新概念对汉字进行剖析,编码
的方法.属全形码类,基本笔画法.米字码可使汉字各种笔画归一化,有序化,象形化排序,
以达到方便记忆,简化,优化编码,降低重码等多种目的.它包括:

- | | |
|-------------------|---------------------|
| 1.笔画分子概念, | B.汉字的米字定律, |
| 2.米字码的三组份, | D.十个数字,四位码,20码元的设定, |
| 3.小键盘米字象形排序法组成代码, | F.千位,百位数部首降重码调控, |
| 4.“数”参与汉字编码, | H.半自动选字法, |
| 5.中心线取码法, | J.米字码引导图. |

笔画分子概念:指任何汉字的各种笔画均可以看成是由一个以上的具有矢量特征的相同的等腰梯形的笔画分子所组成。要点有二:

- (1)任何汉字的笔画均由一个以上相同的等腰梯形的笔画分子所构成。
- (2)等腰梯形的长轴或顶边法线均可赋予矢量标记,故笔画分子在平面空间的角度信息可直接由分子的矢量或其代码表示。

汉字的“米”字定律:“米”字形态具有上下辐射左右对称,45度六等分圆周的特点。全部汉字的笔画分子矢量可以概括为在平面上互差45度的七个不同矢量。可以用米字的七个端点的笔画分子矢量或矢量代码表示,称为汉字的“米字定律”。它使汉字的剖析,编码,取码科学化,使笔画代码间的关系准确有序,并建立起平面形象,符合大脑记忆条件,故好记。此外为简化和节省码符,令“点”的代码只占用一个码符,提出上述“概念,定律”的目的在于:

- (a)简化汉字的基本成份:将其变为归一化的单元即“笔画分子”。
- (b)全部汉字可用此归一化的笔画分子(按平面上总共有七种不同的角度矢量),组合拼成。它提供了形象的,有序的,能与小键盘九键位(自呈米字见图3)良好吻合的笔画代码。
- (c)象形好记不仅提供了取码时判读笔画码的依据,而且可加快学,用。

米字码的三组份有:(a)笔画分子码七个即:[A]横右,[B]竖上,[C]横左(含独横),[D]撇,[E]竖下(含独竖),[F]捺,[G]点,(b)部件码三个即:[H]口,[I]勾,[J]十,X,<。此三含义均称为“交”码,(c)部首码十个即:金,木,水,火,土,月,日,月,日,月,日。

十个数字码符,四位码长和22含义符的设定:米字码采用十个数字码符,四位码长是便于记忆操作两便为出发点,降低重码方法是立足于科学设计:数码分布概率调控及计算机软件辅降重码等措施完成。数字码符的优点:

- (1)比英文字母好记,好编排,易打,简单。
- (2)码符直观有序,见字出码反应快。
- (3)数字码符自身带有“数”的含义,“数”可参与编码,而能起“数码分布概率”重调作用,从而可降低重码,重字,却又不增加码符记忆量负担。

作者所设计的十码符为:0-9十个数字可分别用作笔画分子码,部件码,及部首码共20个码的代码。所选定的十个部首均为汉字的大户部首:金,木,水,火,土等,约占汉字50%,并保留其习惯顺序配置,得出按米字端点矢量旋转排序法的代码即:

[A]为2, [B]3, [C]4, [D]5, [E]6, [F]7, [G]8, [H]0, [I]1, [J]9。
及: 金 61, 木 76, 水 27, 火 37, 土 48, 土 47, 月 67, 日 77, 月 07, 日 17。

小键盘米字象形排序法的代码含义:计算机小键盘的键符共三层,为:1,2,3 4,5,6 7,8,9 其排列正好与米字各端点一一对应:点[7],撇[1],捺[3],竖[5],横[4],此即小键盘米字象形。当采用米字象形排序法编取代码之后,不仅使得七个笔画代码好记,而且空间概念一目了然,易记不怕忘,有助于提高输入速度。以下即为小键盘米字象形排序法得出的代码:

[A]为6, [B]5, [C]4, [D]1, [E]2, [F]3, [G]7, [H]0, [I]8, [J]9
及: 金 23, 木 35, 水 63, 火 53, 土 45, 土 43, 月 23, 日 33, 月 03, 日 83

千位数部首降重码调控:四位数字码可提供十个千位数码,按汉字200个部首计算,平均每千位数下必须容纳10个部首,由于各部首拥有字数不均,导致代码分布严重不均重码多,故采取部首调控措施,令:金,木,水,火,土,月,日,月,日,月,日,十个部首不按首笔笔画,而

特定码,称“降重码调控”,调控前后的汉字分布情况列於表2。由表看出十个特定代码对大降低千位数重码的贡献,它使米字码进入实用成为可能。

表 2. 千位降重码部首变更调控表(小键盘象形米字排序法)

| 千位码 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 总计 (个) |
|-------|----------|-----|----------|-----|-----|-----|------|---------|------|---|-----------|
| 变更部首 | *口 月 | | 𠂇 𠂇 木 | 土 火 | 彳 | 艹 | 点 | *勾 扌 | *交 | | 10 |
| 调前汉字数 | 790 1090 | 260 | 0 | 460 | 320 | 310 | 1400 | 0 | 2130 | | 6760 |
| 调后数 | 740 680 | 620 | 740 | 580 | 680 | 570 | 720 | 710 | 720 | | 6760 |

1. 部首下含字量均为大约数,例:𠂇部(260字),口部(400字),火部(120字),𠂇部(210字),竹部(150字),木部(150字),艹部(330字)。
 2. *号为部件码。由表2.看出经部首调控后,各千位数下拥有汉字量接近平均值(-15% → 10%),对米字码降低重码发生概率,起巨大作用。
- 百位数降重调控:汉字200个部首分配於十个千位码之内,平均每个百位码要接纳两个部首的汉字,而实际情况汉字超过百字大部首很多,百位拥挤需加调控,方法是:
- (a)一般对部首取首尾两码,当部首后续笔画超过4笔时只取首码。(b)选定百位中出现频率较少的“数码3,5”为特定码,性质类似补足码,紧跟特定部首,但非同尾码,不能单独使用。

表 3. 米字码部首码表(双码仅用于后续笔画少于5画) *兼作部件代码

| 千位 | 特定部首码10个 | 部首常规编码(首尾码) 示例 |
|--------------|--|----------------------------------|
| *0 | 月 03 | 口0 日00 目03 田04 里04 黑07 |
| 2 3 | 𠂇 23 𠂇 25 𠂇 33 木 35 | 非22 , 尸 26 26 彳 27 彳 27 隹26 |
| 4 | 土 43 火 45 | 石西酉40 歹页41 耳42 走43 豆工44 瓦47 示豕48 |
| 5 | 彳 53 | 山虫止齿54 业55 虎58 鹵59 |
| 6 | 艹 63 | 骨其60 贝 彭31 马臣64 弓68 牙68 见辰68 风69 |
| *8 | 扌 83 | 羽88 比88 |
| 1 7 *9 | 舌血10 身欠11 爪斤角12 采禾13 鸟乌白14 舟竹17 气毛瓜18 矢父19 户70 母70 广71 𠂇 72 料73 穴75 方76 门衣心78 斗辛羊79 青90 力91 戈92 戈韦丰女93 乡甫94 津96 寸97 豸龙98 皮99 | |

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)